Lakemond, Ruan and Fookes, Clinton B. and Sridharan, Sridha (2009) ***PhD forum : multiple camera management using wide base-line matching.*** In: Third ACM/IEEE International Conference on Distributed Smart Cameras, 30 August - 2 September 2009, Como, Italy.

PhD forum: Multiple Camera Management Using Wide Base-line Matching

Ruan Lakemond, Clinton Fookes, Sridha Sridharan
Image and Video Laboratory, Queensland University of Technology
GPO Box 2434, 2 George St, Brisbane, Queensland 4001
{r.lakemond,c.fookes,s.sridharan}@qut.edu.au

***Abstract*—Manual calibration of large and dynamic networks of cameras is labor intensive and time consuming. This is a strong motivator for the development of automatic calibration methods. Automatic calibration relies on the ability to find correspondences between multiple views of the same scene. If the cameras are sparsely placed, this can be a very difficult task. This PhD project focuses on the further development of uncalibrated wide baseline matching techniques.**

I. Introduction

There exists a vast collection of image processing and computer vision research that is predominantly focused on single camera implementations. This collection is progressively being extended to multiple camera networks. In the past camera networks consisted of sparsely placed cameras with little or no overlap between camera views. The availability of inexpensive sensor networks and the declining price of IP cameras is making it possible to design camera network that have multiple cameras viewing the same scene. Systems of multiple cameras with overlapping views provide far greater possibilities than single camera systems. The most significant advantage is the possibility of full three dimensional scene analysis.

In order for multiple camera networks to deliver more than the sum of many single camera systems, the relationships between cameras must be known and exploited. This requires calibrating the cameras. Methods exist for manually calibrating cameras with high accuracy. Manually calibrating networks with many cameras is, however, time consuming, expensive and unpractical for networks that undergo frequent change. For this reason, automatic calibration techniques have been vigorously researched in recent years.

Fully automatic calibration methods depend on the ability to automatically find point correspondences between overlapping views. Where two cameras are placed close together and have similar focal lengths (short base-line), the two images are very similar and finding correspondences is a simple problem. In typical camera networks cameras are placed far apart in order to cover a larger area. This is referred to as a wide base-line scenario. Finding sufficient correspondences for camera calibration in wide base-line scenarios presents a significant challenge.

This PhD project focuses on developing more effective and efficient techniques for finding correspondences in difficult, uncalibrated, wide base-line multiple camera scenarios. The project consists of two major areas of work. The first is the research and development of more effective and efficient view covariant local feature extractors. The second area of research involves finding methods to extract scene information from a limited set of putative correspondences, in scenarios where local feature matching cannot produce sufficient correspondences to begin calibration.

II. Feature Extraction

Saliency map based local image feature extractors are among the most effective methods for finding correspondences between widely separated views [1], [2], [3], [4]. Although the MSER extractor [5] has shown to be both effective and efficient in wide base-line matching, saliency map based feature extractors provide more features and are more effective when viewing regions that do not consist of predominantly flat surfaces.

Affine covariant saliency map based feature extractors compute a saliency map from an image using a saliency operator such as the Harris cornerness operator or the determinant of Hessian operator. Features are extracted by finding local maxima in the saliency map and then adapting each feature to be scale and affine covariant. The adaptation process is computationally expensive. Several novel, computationally efficient affine adaptation techniques have been developed as part of this PhD program and are presented below.

A. Scale Locus Clustering

A method has been developed to efficiently generate discrete multi-scale representations of local image features. The method operates by clustering a set of multi-scale features according to the scale space locus of each feature. Each resulting cluster represents the evolution over scale of one local feature. The clustering process requires little additional processing after extracting multi-scale features. Since the process operates on the set of multi-scale features only, it does not place any restrictions on the way the scale space pyramid is constructed. This allows greater flexibility and more efficient pyramid computation.

One application of scale locus clustering is a simple and efficient characteristic scale selection algorithm. The Laplacian of Gaussian function can be evaluated along the locus and the maximum point selected as the characteristic scale. The selected point is then at a local maximum in the saliency map and at a local maximum over scales of the Laplacian.

978-1-4244-4620-9/09/$25.00

This method delivers accurate results with little processing and without the need for an iterative technique.

Results from this research has been published in [6].

B. Hessian Matrix Affine Adaptation

Existing affine adaptation methods make use of the second moment matrix to iteratively estimate the local affine shape of local image features. In this research project it has been shown that the Hessian matrix can be used in a similar way to estimate local feature shape. The Hessian matrix can be used to estimate the shape of blob-like structures, but is less effective when used on corner structures.

The Hessian matrix is significantly simpler to compute that the second moment matrix, leading to a significant reduction in computational cost. It may also be more convenient to use the Hessian, for example when used in conjunction with the determinant of Hessian feature extractor.

Results from this research will appear in [7].

III. Uncalibrated Wide Base-line Dense Matching

The first step in the process of automatic calibration and scene reconstruction is usually to find a putative set of correspondences using wide baseline matching. The epipolar geometry is then estimated using the putative correspondences and a robust estimation method, such as RANSAC [8]. The epipolar geometry can then be used to constrain the matching problem and thereby discover more correspondences. In difficult conditions, even the best view covariant feature extraction and matching algorithms may not be able to provide sufficient unique and reliable correspondences to compute an initial estimate of the epipolar geometry. Methods are being researched in this Phd program that will allow the extraction of additional accurate correspondences given only a few initial putative correspondences.

It requires a large amount of information to match a pair of features across widely separated views uniquely. Even when a scene contains a large amount of detailed features, only a small proportion of the features may be unique and of sufficiently large scale to be matched accurately. Each feature that can be matched contains in itself numerous smaller scale features. The smaller features cannot be matched directly, but may be exploited by using the relation between matched features to guide matching. A method called correspondence guided dense matching is under development that can exploit a few accurate matches to extract a large number of fine scale correspondences.

Each feature provides a normalization transform for its surrounding image regions. Normalizing two matched image regions produces a roughly aligned pair of images. Aligning image subregions using features yields inaccurate results because the feature normalization transforms were computed independently from very different images. The images may be of surfaces that are not flat, and therefore it may not be possible to align them perfectly using a linear transform.

The correspondence guided dense matching method proceeds as follows:

1) A pair of matched image regions are carefully aligned to remove the alignment error that results from computing covariant features independently.
2) Points that will be aligned precisely are selected in the aligned images. Points that contain enough fine scale information to avoid the aperture problem are selected.
3) At each selected point, the points are aligned using a small aperture. This allows precise alignment even if the imaged surface is not flat.

Experiments have shown that the correspondence guided dense matching method can produce large numbers of correspondences. The method can produce sufficient matches for computing the epipolar geometry in cases where wide baseline matching has provided insufficient matches.

IV. Summary

This PhD project has delivered more efficient methods for scale selection and affine adaptation of local image features. Currently, methods for extracting large sets of correspondences from difficult wide base-line image pairs are under development

References

[1] A. Baumberg, “Reliable feature matching across widely separated views,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, Hilton Head Island, South Carolina, USA, 2000, pp. 774–781.

[2] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, pp. 63–86, 2004.

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[4] P. Moreels and P. Perona, “Evaluation of features detectors and descriptors based on 3d objects,” *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[5] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” in *British Machine Vision Conference, BMVC 2002*, P. R. a. D.Marshall, Ed., vol. 1, Cardiff, UK, 2002, pp. 384–393.

[6] R. Lakemond, D. McKinnon, C. Fookes, and S. Sridharan, “A feature clustering algorithm for scale-space analysis of image structures,” in *International Conference on Signal Processing and Communication Systems, ICSPCS'2007*, Gold Coast, Australia, 2007.

[7] R. Lakemond, C. Fookes, and S. Sridharan, “Affine adaptation of local image features using the hessian matrix,” in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, to appear, 2009.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York: Cambridge University Press, 2003.